

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-257790

(43)Date of publication of application : 03.10.1997

(51)Int.Cl.

G01N 33/48

G01N 33/50

G01N 33/66

(21)Application number : 09-001921

(71)Applicant : NAKANO VINEGAR CO LTD
IWATA AKIRA

(22)Date of filing : 09.01.1997

(72)Inventor : YAMADA EMIKO
NAKAGAWA HIROAKI
TAKAHASHI REIKO
TSUKAMOTO YOSHINORI
KAWAMURA KICHIYA
OGURI KOJI
IWATA AKIRA

(30)Priority

Priority number : 08 7426

Priority date : 19.01.1996

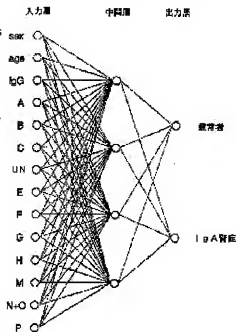
Priority country : JP

(54) DIAGNOSIS SUPPORTING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To support the diagnosis of a disease and the others with high accuracy by using the content of a plurality of sugar chains contained in glycoprotein or glycolipid in tissue and a tissue soln.

SOLUTION: Data A, B,...P correlated with the concns. of a plurality of sugar chains sampled from a living body are processed by a neutral network to obtain data useful for diagnosis. The data correlated with the concns. of sugar chains are the data directly or indirectly correlated with the concns. of sugar chains such as the concns. or concn. ratio of sugar chains, the areas or areal ratio of peaks detected by HPLC or the like. Clinical data necessary for diagnosis such as an age or the distinction of sex or the like other than said data may be simultaneously inputted to the neutral network. When sugar chains are ones obtained from IgG, the diagnosis of IgA nephropathy, rheumatoid, atopic dermatitis, diabetes and liver diseases can be supported.



*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1]A diagnosis support method processing data correlated with concentration of two or more sugar chains extracted from a living body by a neural network, and obtaining a diagnostic-assistances result.

[Claim 2]The diagnosis support method according to claim 1 processing simultaneously clinical data other than data correlated with concentration of said sugar chain by said neural network.

[Claim 3]The diagnosis support method according to claim 1 or 2 with which data correlated with concentration of said sugar chain is characterized by missing data and flying apart, assigning a value for exclusive use respectively to a value, and inputting into said neural network.

[Claim 4]When data correlated with concentration of said sugar chain makes a and b a number which fills $0 < a < b < 1$, The diagnosis support method according to claim 1 or 2 which sets missing data to 0 and is characterized by flying apart, setting a value to 1, normalizing other data so that it may be distributed between a and b, and inputting into said neural network.

[Claim 5]A diagnosis support method of claim 1-4 said sugar chain's being a sugar chain obtained from IgG, and performing diagnosis of IgA glomerulonephritis or rheumatism given in any 1 paragraph.

[Claim 6]A diagnosis support method of claim 1-4 said sugar chain's being a sugar chain obtained from IgG, and diagnosing liver disease which consists of atopic dermatitis, diabetes mellitus or hepatitis, liver cirrhosis, or hepatoma given in any 1 paragraph.

[Claim 7]A diagnosis support method of claim 1-4 said sugar chain's being a sugar chain obtained from IgG, and identifying severity or typus morbi of a disease given in any 1 paragraph.

[Claim 8]A diagnosis support method of claim 1-4 said sugar chain's being a sugar chain obtained from IgG, and identifying a monopathy from between two or more diseases given in any 1 paragraph.

[Translation done.]

*** NOTICES ***

JP0 and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the diagnosis support method which performs diagnostic assistances, such as an illness, based on clinical data.

[0002]

[Description of the Prior Art] Much glycoprotein and a glycolipid are in a human organization and tissue fluid, and dozens of kinds of sugar chains contained in them exist. For example, about 40 kinds of N knot-pattern sugar chains are contained in glycoprotein in a blood serum (Nakagawa, H., et al. (1995), Anal. Biochem. 226, 130-138.). It is known that, as for a neutral sugar chain of the immunoglobulin G of a blood serum, there are varieties (Mizuochi, T., et al. (1982), J. Immunol., 129, 2016.). Methods of measuring the content of these sugar chains include the method of using high performance chromatography (HPLC), the method of using lectin, an antibody, etc., the method of using immunoelectrophoresis, etc.

[0003] It is known that two or more sugar chain structures included in glycoprotein in an organization and tissue fluid and a glycolipid are changing with various diseases now. For example, in rheumatism. It is known that sugar chain concentration of several kinds is changing among 16 kinds of sugar chain structures of the immunoglobulin G (IgG) in a blood serum (Springer Semin. Thomas W. Rademacher, et al. (1988)). Immunopathol., 10, and 231-249.. In the hepatoma, it turns out that the sugar chain structure included in the transferrin in a blood serum is changing (Yamashita, K., et al. (1989), J.Biol. Chem., 264, 2415-2423.). From cancer, the sugar chain structure of a cell cortex changes and the monoclonal antibody which uses the sugar chain as an antigen is actually used already as a tumor marker. In a pancreatic cancer, the carbohydrate antigen is a tumor marker according to the high positive rate of 80% (Umeyama et al.: the pancreas, 3, 22-33, 1988).

[0004]

[Problem(s) to be Solved by the Invention] The measured value of the single sugar chain content using the monoclonal antibody which uses a sugar chain as an antigen is only put in practical use as a tumor marker at present, and the disease diagnostic assistances by the measured value of two or more sugar chain content which used other measuring methods, such as detection system by HPLC, for example are not put in practical use.

[0005] In order to use sugar chain content for disease diagnosis, it is necessary to measure the sugar chain content of the organization of a patient and a healthy person and glycoprotein in tissue fluid, and a glycolipid, and to measure the sugar chain content of a healthy person and a patient. Usually, this comparison is performed by the statistical technique of calculating a patient's positive rate by setting up a cutoff value.

[0006] In this statistical technique, it analyzes one kind of sugar chain content at a time altogether for every kind of sugar chain, and a healthy person is compared with a patient combining the content of several kinds of sugar chains, such as a sugar chain with whether a healthy person and a patient compare each and by sect N-acetyl glucosamine, and a sugar chain which it does not have. Since what has the difference and sex difference by age in the content exists in a sugar chain in the living body, they must also be taken into consideration,

must perform comparison with a patient and a healthy person synthetically, and must use it for disease diagnosis. However, it is difficult for age, sex, etc. to conduct sufficient analysis in the present statistical technique, in consideration of the top where operation of the combination of the kind of sugar chain, setting out of a cutoff value, etc. is complicated, since there are many kinds of sugar chain.

[0007] Thus, although it is clear that a sugar chain has a disease etc. and correlation, in the statistical technique used now, there is a problem that the synthetic analysis of two or more sugar chain content contained in glycoprotein in an organization and tissue fluid and a glycolipid is difficult.

[0008] This invention was made in view of such the actual condition, and an object of this invention is to provide the method that a high-precision disease and others can be diagnosed using two or more sugar chain content contained in glycoprotein and the glycolipid in an organization and tissue fluid.

[0009]

[Means for Solving the Problem] In this invention, said purpose is attained by processing data correlated with concentration of two or more sugar chains extracted from a living body by a neural network, and acquiring useful information to diagnosis.

[0010] With data correlated with concentration of a sugar chain, for example Concentration or the ratio of concentration of a sugar chain. Data directly or indirectly relevant to concentration of a sugar chain, such as dyeing concentration detected by intensity, such as condensation, coloring, and luminescence, and immunoelectrophoresis which were detected by area or surface ratio, lectin, and an antibody of a peak detected by HPLC, is meant.

[0011] Clinical data, such as the other age required for diagnosis, sex, a white blood cell count (WBC), and IgG concentration, is suitably chosen as a neural network, and it may be made to input it into him simultaneously with data correlated with concentration of a sugar chain.

[0012] It is preferred to assign a value for exclusive use respectively to those data, and to input into data correlated with concentration of a sugar chain at a neural network, missing data or when it flies apart and there is a value (it compares with other values and is data of an extremely big value or a small value). for example, a time of data correlated with concentration of a sugar chain making a and b a number which fills $0 < a < b < 1$ -- missing data -- $- 0$ -- it flies apart and a value is set to 1, and it normalizes so that it may be distributed between a and b, and other data is inputted into said neural network. When an example of a value of a and b is shown, it is $a = 0.1$ and $b = 0.9$.

[0013] If data correlated with concentration of a sugar chain is used as sugar chain data obtained from IgG, IgA glomerulonephritis, rheumatoid arthritis, diabetes mellitus, liver disease, or information useful to diagnosis of atopic dermatitis can be acquired.

[0014] A neural network has high-speed parallel operation capability and a learning function, and demonstrates power to analysis of much information. As a neural network, can adopt a layered network, a cross coupling type network, etc. suitably, for example, and as the learning method, An error reverse propagation learning method etc. of common knowledge which changes coupling load and determines optimal coupling load so that an error of an output pattern given as a teacher pattern and an output value of neurone of a actual output layer may be decreased one by one are employable.

[0015]

[Embodiment of the Invention] Hereafter, an embodiment of the invention is described.

[Embodiment 1] The discernment examination of IgA glomerulonephritis was done using the data of the IgG neutral sugar chain of diagnostic-assistances plurality of IgA glomerulonephritis. IgG neutral sugar chain structure Age (Raj. Parekh, et al. (1988), Brief Definitive Report, 167, 1731-1736.), Since changing with sex (N. Tsuchiya, et al. (1993), J.Immun., 151, 1137-1146.) is known, The data of the peak area ratio (percentage) of age, sex, IgG concentration, and an IgG neutral sugar chain was used as an input pattern. The thing of 19 IgA glomerulonephritis patients and 30 healthy persons with confirmed diagnosis was used for the data for a discernment examination.

[0016] The IgG neutral sugar chain was analyzed as follows. First, IgG was prepared using the

protein G from the blood serum. While cutting down the sugar chain by enzyme digestion from prepared IgG, all the sugar chains were used as the neutral sugar chain by enzyme digestion. The cut-down neutral sugar chain was fluorescent-labeling-ized, and HPLC using an octadecylsilyl (it is described as octadecylsilyl and the following ODS) column (Nakanopak ODS-A: Product made from AJINOKI) analyzed 12 kinds of neutral sugar chains. The sugar chain structure of the analysis result by an ODS column and each peak A-P is shown in drawing 1.

[0017]When calculating a peak area ratio (percentage), since the sugar chain of the serum glycoprotein origin of those other than IgG may have mixed irrespective of a IgA glomerulonephritis patient and a healthy person by a sample, the peak D was excepted from input data. Although the peak N and the peak O were originally divided into two peaks, since it may have been detected as one peak by a column, it calculated as one peak of N+O beforehand. It asked for the percentage of each peak area from the area of 11 kinds of all sugar chain peaks detected other than the peak D.

[0018]First, it examined identifying IgA glomerulonephritis by the test-of-significance method of conventional statistical Mann-Whitney using the neutral sugar chain peak area ratio of IgG calculated as mentioned above for comparison. As a result, the significant difference was observed only in the ratio of the sugar chain peaks F and G shown in drawing 2 by 5% or less of percentage of risk. As for the recognition rate of the IgA glomerulonephritis patient at the time of setting a cutoff value as average value \pm (2x standard deviation), and a healthy person, at this time, the specificity was 10.5% in sensitivity 100% 65.3%. However, a recognition rate, specificity, and sensitivity are defined by the following formula, respectively, when setting to nh the number of samples which identified np and a healthy person for the number of samples which identified [the number of patient samples] NH and a patient for NP and the number of healthy person samples as a patient as a healthy person.

[0019]Recognition rate = $(np+nh)/(NP+NH)$

Sensitivity = np/NP specificity = nh/NH [0020]Next, the method of this invention using a neural network is explained. Here, it asked for the percentage of each peak area from the area of 11 kinds of all sugar chain peaks detected other than the peak D shown in drawing 1, and was considered as input data. Having decided to use all the detected peaks as an input pattern is based on the following reasons. The sugar chain structure said to change with diseases until now is caught only by change of remarkable concentration, and change of the statistic by the combination of the simple sugar chain on statistics. However, considering that many kinds of sugar chains exist in the living body, in the analysis conducted until now, since a possibility that the information which a sugar chain has cannot fully be pulled out is left behind, analysis about all the sugar chains will be conducted by the complicated pattern analysis by a neural network.

[0021]Among [A, B, and C] the sugar chain peaks [peaks / said / 11 kinds of / sugar chain] further shown in drawing 1, except for four kinds of data of UN, it can ask for the percentage of each sugar chain peak area from remaining seven kinds, and can use as input data. Since the reaction by-product produced in the case of fluorescent-labeling-izing of a sugar chain laps with the elution position from which the sugar chain peak A, B, and C and UN are detected by HPLC using an ODS column and detected, if the sugar chain peak A, B, and C and UN are excluded, dispersion by the reaction condition of data can be held down to it.

[0022]When the discernment examination of a monopathy was done from two or more diseases, using the peak area ratio of seven kinds of such IgG neutral sugar chains as input data, sensitivity and a recognition rate improved slightly. As a neural network, the layered network of three layers as shown in drawing 3 was adopted.

[0023]Here, as an input pattern, age, sex, IgG concentration, and the data of the peak area ratio (percentage) of 11 kinds of said IgG neutral sugar chains, That is, since it identified in all whether they were two kinds of outputs, i.e., IgA glomerulonephritis, using 14 items of data, the number of neurone of the input layer made two pieces 14 pieces and the number of neurone of the output layer. Two neurone of an output layer supports the output values 0 (he

is a healthy person) and 1 (it is IgA glomerulonephritis). It was considered as the number of neurone of the output layer with two pieces in this example in order to only clarify significance of discernment, and the number of neurone of an output layer may be one piece, and it may be made to display the probability or percentage of risk of a possibility of being IgA glomerulonephritis as three or more pieces.

[0024] If too small, a complicated combination of an input pattern will not get across to an output layer, and it will not be completed by study, and even if there is too many an interlayer's neurone conversely, its learning frequency increases, it requires time and becomes complicated. Therefore, the interlayer's number of neurone determined the optimum number experimentally. Here, changed the interlayer's number of neurone from 1 to 30, it was made to learn by the error reverse spreading method, and the number of neurone of the interlayer at the time of being the highest-achieving and four pieces were adopted in having conducted the discernment experiment.

[0025] 30 IgA glomerulonephritis patients who a neural network's study took out 11 examples from 30 healthy persons' data with the above-mentioned confirmed diagnosis, and 19 IgA glomerulonephritis patients' data at random, and considered it as 30 examples. The output pattern to wish to have was twisted to the error reverse propagation learning method given as a teacher pattern using the data of a total of 60 examples. First, each of each peak area ratio of 11 kinds of IgG neutral sugar chains by age, IgG concentration, and the above-mentioned analysis method was normalized by the following (several 1). However, it was considered as the neural network who can cope with lack information in inputting as $a = 0$ to missing data, and was referred to as $a = 1$ to extremely big data or small data (flying apart value) compared with other values. About sex, 0.1 and a woman were evaluated for the male as 0.9, the value which flies apart and can reject a value, for example by Smirnov's null hypothesis test — carrying out (the 284-285th pages of Kiyoshi Ichihara "statistics of bioscience" Nankodo publication (1990)) — or it may be decided experientially that they will be arbitrary values. Here, all values rejectable by Smirnov's null hypothesis test were set to $a = 1$, it normalized by (several 1), and the unrejectable value was evaluated and inputted between 0.1-0.9.

[0026]

[Equation 1] $\text{Data } a_{\text{max}} : \text{maximum } a_{\text{min}} \text{ before the data } a_0; \text{normalization after } a = (a_0 - a_{\text{min}}) / (a_{\text{max}} - a_{\text{min}}) \times 0.8 + 0.1; \text{normalization}$ [0027] An input value multiplies by coupling load, respectively, and gets across to an interlayer, and each input value by which coupling load multiplied is added, and is outputted in response to modification by a response function. Coupling load takes the advantage of the output from an interlayer again, and it is added to it in an output layer, and is outputted in response to modification by a response function. The sigmoid function was used as a response function. Coupling load is changed to the optimal value by repeating changing coupling load so that the error of the given teacher pattern and the value of a actual output layer may be decreased one by one to all the teacher patterns with an error reverse propagation learning method.

[0028] The above mentioned data of a IgA glomerulonephritis patient and 30 healthy persons each (a total of 60 examples) is used, The reeve one out (leave one out) method, Namely, the place which investigated the IgA glomerulonephritis patient's recognition rate and sensitivity by the method of repeating respectively identifying about the one remaining examples using the neural network who did the end of study by the data of 59 examples, and performing it about all the 60 examples data, It compares with the conventional statistical technique which the recognition rate explained at 93.3% and sensitivity explained according to 96.7% and a comparative example, and the gap also improved by leaps and bounds.

[0029] Next, each sugar chain data found out as follows how it would contribute to discernment of IgA glomerulonephritis. As shown in drawing 4, change of a recognition rate when combination with one arbitrary input layer (an example of drawing 4 the peak P) and interlayer is intercepted among 11 kinds of IgG neutral sugar chain peak area ratios was investigated about 11 kinds of all the sugar chains. When setting a recognition rate when

combination with an entry-of-data layer of x_0 and one kind of a certain sugar chain peak area ratio and an interlayer is intercepted for a recognition rate at the time of combining all of an entry-of-data layer and an interlayer of a sugar chain peak area ratio to x . The following (several 2) defines the contribution y over a recognition rate of a sugar chain which intercepted combination with the input layer and interlayer.

[0030]

[Equation 2] $y = (x_0 - x) / x_0$ [0031] Contribution is similarly defined about specificity and sensitivity. To each of a recognition rate, specificity, and sensitivity, drawing 5 graph-izes the contribution of each sugar chain data in a IgA glomerulonephritis patient's discernment, and shows it. The useful thing is meant to discernment of the IgA glomerulonephritis by a neural network as the sugar chain in which big contribution is shown. In this case, it turns out that all the sugar chain data other than the sugar chain peak P is acting effectively about discernment of IgA glomerulonephritis, a neural network discriminates from drawing 5 with a complex combination of these sugar chain data, and the recognition rate is raised.

[0032] A discriminated result by the conventional statistical technique and a discriminated result by a neural network who combined all the neurone are summarized in Table 1, and are shown.

[0033]

[Table 1]

1 IgA腎症の識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	10.5	100	65.3
ニューラルネットワーク (本発明)	88.7	99.0	98.3

[0034] Thus, by using a neural network as a means to analyze IgG neutral sugar chain peak area ratio data, Analysis which combined sugar chain information on various sorts which were not fully able to be performed by the conventional statistical technique can be conducted automatically, and, moreover, improvement in a recognition rate of a disease and a non-disease in the analysis can be aimed at.

[0035] Although HPLC which used an ODS column analyzed a sugar chain and an example which adopts data of each peak area ratio as an input pattern was explained here. With a method, immunoelectrophoresis, etc. which use other columns (gel filtration, an affinity column, etc.), or use lectin and an antibody, detection of a sugar chain is good and may use data of concentration of a sugar chain measured by these methods, etc. as an input pattern.

[0036] [Embodiment 2] A discernment examination of rheumatoid arthritis was done using data of a peak area ratio (percentage) of an IgG neutral sugar chain analyzed by HPLC using diagnostic-assistances age of rheumatoid arthritis, sex, IgG concentration, the number of WBC (leucocytes), and an ODS column as an input pattern. As data for a discernment examination, a thing of 59 sufferers from rheumatism and 278 healthy persons with confirmed diagnosis was used.

[0037] Analysis of an IgG neutral sugar chain was conducted like Embodiment 1. The sugar chain structure of the analysis result by an ODS column and each peak A-P is shown in drawing 6. It examined first, using the neutral sugar chain peak area ratio of IgG, and identifying rheumatoid arthritis by the test-of-significance method of conventional statistical Mann-Whitney for comparison. As a result, as shown in drawing 7, as for the recognition rate of the sufferer from rheumatism and healthy person at the time of setting a cutoff value as average value ± 2 (2x standard deviation) by the ratio of the sugar chain peaks H and P the significant differences were accepted to be by 0.1% or less of percentage of risk, the specificity was 66.1% in sensitivity 96.8% 91.4%.

[0038] Next, the identifying method using a neural network is explained. The neural network was taken as the layered network of three layers using the same sigmoid function as Embodiment 1, as shown in drawing 8. Since it identified whether it was rheumatism based on the information on [15 items of / a sugar chain and clinical], the number of neurone of the

input layer was made into 15 pieces, and the output layer could be two pieces. Two neurone of an output layer is the output values 0 (it is healthy) and 1 (it is rheumatism). It corresponds. The interlayer's number of neurone determined the optimum number by experiment. Here, the number of neurone of the interlayer at the time of changing an interlayer's number of neurone to 1-30, making it learn by the error reverse spreading method, and being able to learn most efficiently was adopted. In this case, they could be 16 pieces.

[0039]A neural network's study was based on the error reverse propagation learning method which gives the output pattern to wish to have as a teacher pattern using the data of a rheumatoid arthritis patient and 50 healthy persons each with the above-mentioned confirmed diagnosis. The network which reached shortest time was used for the set-up error rate supposing the case where there is practical restriction this time. First, the normalizing process by the above (several 1) was performed to each of the sugar chain peak area ratio of 11 kinds of IgG neutral sugar chains by age, IgG concentration, the number of WBC, and the above-mentioned analysis method like Embodiment 1, and it was considered as the input pattern. However, to missing data, it was considered as the network which can cope with lack information as $a=0$, and was referred to as $a=1$ to extremely big data or small data (flying apart value) compared with other values. About sex, 0.1 and a woman were evaluated for the male as 0.9.

[0040]The above mentioned data of a rheumatoid arthritis patient and 50 healthy persons each (a total of 100 examples) is used, The reeve one out (leave one out) method, Namely, the place which investigated the rheumatoid arthritis patient's recognition rate by the method of repeating respectively identifying about the one remaining examples using the neural network who did the end of study by the data of 99 examples, and performing it about all the 100 examples data, It became 96.0% of a recognition rate, 96.0% of specificity, and 96.0% of sensitivity, and sensitivity improved.

[0041]Next, it was found out how each sugar chain data would contribute to discernment of rheumatoid arthritis like Embodiment 1. Like drawing 5, drawing 9 graph-izes the contribution of each sugar chain in a rheumatoid arthritis patient's discernment, and shows it. The larger sugar chain of contribution is more useful to discernment of the rheumatoid arthritis patient by a neural network.

[0042]When combination with the input layer of the data of the sugar chain peak B and an interlayer was intercepted as data of the low sugar chain of contribution to discernment of the rheumatoid arthritis patient by a neural network with reference to the result of drawing 9, specificity improved to 98.0% and the recognition rate improved to 97.0%.

[0043]The discriminated result by the neural network at the time of combining all of the discriminated result by the conventional statistical technique, and the entry-of-data layer and interlayer of a sugar chain and the discriminated result by the neural network who intercepted combination with the entry-of-data layer of the sugar chain peak B and an interlayer are summarized in Table 2, and are shown.

[0044]

[Table 2]

慢性関節リウマチの鑑別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	66.1	96.8	91.4
全てのニューロを結合した ニューラルネットワーク (本発明)	96.0	96.0	96.0
貢獻度の低いニューロとの 結合を遮断したニューラル ネットワーク (本発明)	96.0	98.0	97.0

[0045][Embodiment 3] The discernment examination of atopic dermatitis was done using the data of the peak area ratio (percentage) of the IgG neutral sugar chain analyzed by HPLC

using the diagnostic-assistances age of atopic dermatitis, sex, and an ODS column as an input pattern. As data for a discernment examination, the thing of 24 atopic dermatitis patients and 63 healthy persons with confirmed diagnosis was used.

[0046]First, when it examined identifying atopic dermatitis by the test-of-significance method of conventional statistical Mann-Whitney for comparison, as shown in drawing 10, the significant difference was observed in the ratio of the sugar chain peaks F and G by 5% or less of percentage of risk. As for the recognition rate of the atopic dermatitis patient at the time of setting a cutoff value as average value \pm (2x standard deviation), and a healthy person, at this time, the specificity was 8.3% in sensitivity 92.1% 69.0%.

[0047]Next, the identifying method using a neural network is explained. The number of neurone of a neural network's input layer was made into 13 of age, sex, and 11 kinds of IgG sugar chains, and the number of neurone of the output layer was made into atopic dermatitis and two healthy pieces. The interlayer's number of neurone was determined based on the rule of thumb, and was made into four pieces. Each data of age, sex, and an IgG sugar chain performed the normalizing process like Embodiment 1, and used it as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern wished to have using the data of 24 examples selected at random out of 24 atopic dermatitis patients with the above-mentioned confirmed diagnosis, and 63 healthy persons as a teacher pattern.

[0048]When the atopic dermatitis patient's recognition rate was investigated by the reeve one out method using the data of this atopic dermatitis patient and 24 healthy persons each (a total of 48 examples), 89.4% of sensitivity, 83.3% of specificity, and 86.3% of a recognition rate were acquired. Thus, by using a neural network, sensitivity and a recognition rate improved compared with the conventional statistical technique.

[0049]The discriminated result by the conventional statistical technique and the discriminated result by a neural network are summarized in Table 3, and are shown.

[0050]

[Table 3]

アトピー性皮膚炎の識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	8.3	92.1	69.0
ニューラルネットワーク (本発明)	89.4	83.3	86.3

[0051][Embodiment 4] The diabetic discernment examination was done using the data of the peak area ratio (percentage) of the IgG neutral sugar chain analyzed by HPLC using diabetic diagnostic-assistances age, sex, and an ODS column as an input pattern. As data for a discernment examination, the thing of 34 diabetics and 110 healthy persons with confirmed diagnosis was used.

[0052]First, when it examined identifying diabetes mellitus by the test-of-significance method of conventional statistical Mann-Whitney for comparison, as shown in drawing 11, the significant difference was observed in the rate of the sugar chain peak F by 0.1% or less of percentage of risk. As for the recognition rate of the diabetic at the time of setting a cutoff value as average value \pm (2x standard deviation), and a healthy person, at this time, the specificity was 2.9% in sensitivity 95.5% 73.6%.

[0053]Next, the identifying method using a neural network is explained. The number of neurone of a neural network's input layer was made into 13 of age, sex, and 11 kinds of IgG sugar chains, and the number of neurone of the output layer was made into diabetes mellitus and two healthy pieces. The interlayer's number of neurone was determined based on the rule of thumb, and was made into four pieces. Each data of age, sex, and an IgG sugar chain performed the normalizing process like Embodiment 1, and used it as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern wished to have using the data of 34 examples selected at random out of 34 diabetics with the above-mentioned confirmed diagnosis, and 110 healthy persons

as a teacher pattern.

[0054]When the diabetic's recognition rate was investigated by the reeve one out method using the data of this diabetic and 34 healthy persons each (a total of 68 examples), 100% of sensitivity, 100% of specificity, and 100% of a recognition rate were acquired. Thus, by using a neural network, sensitivity, specificity, and a recognition rate improved compared with the conventional statistical technique.

[0055]The discriminated result by the conventional statistical technique and the discriminated result by a neural network are summarized in Table 4, and are shown.

[0056]

[Table 4]

糖尿病の識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	2.9	95.5	73.6
ニューラルネットワーク (本発明)	100	100	100

[0057][Embodiment 5] The discernment examination of liver disease was done using the data of the peak area ratio (percentage) of the IgG neutral sugar chain analyzed by HPLC using the diagnostic-assistances age of liver disease, sex, and an ODS column as an input pattern. As data for a discernment examination, the thing of 60 patients with liver disease (20 hepatitis, 20 liver cirrhosis, 20 hepatoma) and 110 healthy persons with confirmed diagnosis was used.

[0058]First, when it examines identifying liver disease by the test-of-significance method of conventional statistical Mann-Whitney for comparison, as shown in drawing 12, the ratio of a sugar chain with by sect N-acetyl glucosamine to the sugars which it does not have -- the significant difference was observed in (M+N+O+P)/(E+F+G+H) by 0.1% or less of percentage of risk. As for the recognition rate of the patient with liver disease at the time of setting a cutoff value as average value ** (2x standard deviation), and a healthy person, at this time, the specificity was 43.3% in sensitivity 93.6% 75.9%.

[0059]Next, the identifying method using a neural network is explained. The number of neurone of a neural network's input layer was made into 13 of age, sex, and 11 kinds of IgG sugar chains, and the number of neurone of the output layer was made into liver disease and two healthy pieces. The interlayer's number of neurone was determined based on the rule of thumb, and was made into four pieces. Each data of age, sex, and an IgG sugar chain performed the normalizing process like Embodiment 1, and used it as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern wished to have using the data of 60 examples selected at random out of 60 patients with liver disease with the above-mentioned confirmed diagnosis, and 110 healthy persons as a teacher pattern.

[0060]When the patient's with liver disease recognition rate was investigated by the reeve one out method using the data of this diabetic and 60 healthy persons each (a total of 120 examples), 87.8% of sensitivity, 83.9% of specificity, and 85.9% of a recognition rate were acquired. Thus, by using a neural network, sensitivity and a recognition rate improved compared with the conventional statistical technique.

[0061]The discriminated result by the conventional statistical technique and the discriminated result by a neural network are summarized in Table 5, and are shown.

[0062]

[Table 5]

肝疾患の識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	43.3	93.6	75.9
ニューラルネットワーク (本発明)	87.8	83.9	85.9

[0063][Embodiment 6] Diagnostic assistances of rheumatoid arthritis (severity discernment) Severity was identified by the neural network using the data of a rheumatoid arthritis patient and a healthy person. The data used as an input pattern is the same as Embodiment 2.

[0064]As data for a discernment examination, 48 rheumatoid arthritis patients and 282 healthy persons of confirmed diagnosis who are attached were used. The rheumatoid arthritis patient is divided into the stages 1-4 as severity based on organic change of a joint. In now, in the patient of the stage 1, the patient of 15 examples and the stage 2 added the data for which the patient of 11 examples and the stage 3 chose for the number of the patients of six examples and the stage 4 to be 16 at random to make each 20 stages. The classification of severity is based on the stage classification of Steinbrocker (this country clinical statistics collection and 512-521 (1993) Nippon Rinsho). The healthy person chose 20 examples from 282 examples at random.

[0065]The number of neurone of a neural network's input layer was made into 15 pieces like Embodiment 2, and the number of neurone of the output layer was made into each stage and five healthy pieces. The interlayer's number of neurone was determined based on the rule of thumb, and was made into six pieces. Each data of age, sex, IgG concentration, the number of WBC, and an IgG sugar chain performed the normalizing process like Embodiment 1, and used it as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern wished to have using the data of 20 examples selected at random out of 20 rheumatoid arthritis patients each of the stage 1 - the stage 4 with the above-mentioned confirmed diagnosis, and 282 healthy persons as a teacher pattern.

[0066]When the severity of rheumatoid arthritis is identified by the reeve one out method using the data of this rheumatoid arthritis patient and a healthy person, as shown in Table 6, 84.5%, 97.5%, the stage 3 was 100.0%, the stage 4 was 87.5%, and the stage 1 of sensitivity was [the stage 2] an average of 92.4%. 96.0% of specificity and 93.1% of a recognition rate were acquired. Thus, the severity of rheumatoid arthritis was able to be identified by using a neural network.

[0067]

[Table 6]

慢性関節リウマチの重症度識別結果

平 均	感 度 (%)				特異度 (%)	認識率 (%)
	ステージ1	ステージ2	ステージ3	ステージ4		
92.4	84.5	97.5	100.0	87.5	96.0	93.1

[0068][Embodiment 7] Diabetic diagnostic assistances (typus morbi discernment)

The neural network performed diabetic typus morbi discernment using the data of a diabetic and a healthy person. As an input pattern, the data of age, sex, and an IgG sugar chain was used.

[0069]As data for a discernment examination, 34 diabetics (inside five I-beams, II type 29 examples) and 110 healthy persons with confirmed diagnosis were used. The classification of diabetic typus morbi is based on a classification of WHO in 1980 (internal medicine, 25, 1524 - 1527 (1995) Nankodo). What has shortage of type I diabetes mellitus and an insulin relative [what has absolute shortage of an insulin] is classified into type II diabetes mellitus. Type I diabetes mellitus also calls insulin-dependent diabetes mellitus and type II diabetes mellitus non-insulin dependent diabetes mellitus. The patient added data selected at random and was taken each as an I-beam and 30 II molds. The healthy person chose 40 examples from 110 examples at random.

[0070]The number of neurone of a neural network's input layer was made into 13 of age, sex, and 11 kinds of IgG sugar chains, and the number of neurone of the output layer was made into a diabetes-mellitus I-beam, a diabetes-mellitus II type, and three healthy pieces. The interlayer's number of neurone was determined based on the rule of thumb, and was made into four pieces. Each data performed the normalizing process like Embodiment 1, and used it

as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern to wish to have as a teacher pattern.

[0071]When diabetic typus morbi was identified by the reeve one out method using the data of this diabetic and a healthy person, as shown in Table 7, a diabetes-mellitus I-beam is [100.0% and a diabetes-mellitus II type] 93.2%, and, as for sensitivity, an average of 96.7% was obtained. 100.0% of specificity and 98.0% of a recognition rate were acquired. Thus, diabetic typus morbi was able to be identified by using a neural network.

[0072]

[Table 7]

糖尿病の病型識別結果

平均	感 度 (%)		特異度 (%)	認識率 (%)
	I 型	II 型		
96.7	100.0	93.2	100.0	98.0

[0073][Embodiment 8] Two or more monopathies were discriminated from the disease by the neural network using the data of the patient for discernment each disease of every of a monopathy, and a healthy person's data. As an input pattern, the data of age, sex, and an IgG sugar chain was used.

[0074]19 IgA glomerulonephritis patients and 30 healthy persons who attach confirmed diagnosis as data for a discernment examination, respectively, 48 rheumatoid arthritis patients, 282 healthy persons, 24 atopic dermatitis patients and 63 healthy persons. Take out 30 examples and a healthy person six examples at a time from 34 diabetics, 110 healthy persons, 60 patients with liver disease and 110 healthy persons, and the above data at random, respectively, and a patient on the whole. The discernment examination was done as a total of 180 examples of 30 IgA glomerulonephritis patients, 30 rheumatoid arthritis patients, 30 atopic dermatitis patients, 30 diabetics, 30 patients with liver disease, and 30 healthy persons. However, about the atopic dermatitis patient and the IgA glomerulonephritis patient, the data selected at random out of each patient was added to make 30 examples.

[0075]The number of neurone of a neural network's input layer was made into 13 of age, sex, and 11 kinds of IgG sugar chains, and the number of neurone of the output layer was made into six pieces, a IgA glomerulonephritis patient, a rheumatoid arthritis patient, an atopic dermatitis patient, a diabetic, a patient with liver disease, and a healthy person. The interlayer's number of neurone was determined based on the rule of thumb, and was made into 12 pieces. Each data performed the normalizing process like Embodiment 1, and used it as the input pattern. A neural network's study was based on the error reverse propagation learning method which gives the output pattern to wish to have as a teacher pattern.

[0076]When a disease is identified by the reeve one out method using the data of said two or more disease patient and a healthy person, as shown in Table 8, Sensitivity was IgA glomerulonephritis, and of rheumatism, 81.7%, 84.3%, 86.3% was obtained of diabetes mellitus by atopic dermatitis, and it was obtained on the average 88.0% by liver disease 87.0% 90.7%. Specificity was 61.3% and the recognition rate was 82.2%. Thus, by using a neural network, the monopathy was able to be discriminated from two or more diseases.

[0077]

[Table 8]

複数の疾患から単一疾患の識別結果

平均	感 度 (%)					特異度 (%)	認識率 (%)
	IgA腎症	リウマチ	アトピー	糖尿病	肝疾患		
86.3	90.7	81.7	84.3	87.0	88.0	61.3	82.2

[0078]

[Effect of the Invention]According to this invention, analysis which combined the sugar chain content of the various sorts which were not fully able to be performed by the conventional

statistical technique can be automatically conducted by using a neural network as a means to analyze sugar chain data.

[Translation done.]

*** NOTICES ***

JPO and INPIT are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1]The figure showing a IgA glomerulonephritis patient's IgG neutral sugar chain analysis result.

[Drawing 2]The distribution map of sugar chain peak ratio G/F to IgA glomerulonephritis and a healthy person.

[Drawing 3]The lineblock diagram of the neural network for IgA glomerulonephritis patient discernment.

[Drawing 4]The explanatory view of the neural network who intercepted combination with an input layer and an interlayer about one sugar chain data.

[Drawing 5]The figure showing the contribution of each sugar chain data in a IgA glomerulonephritis patient's discernment.

[Drawing 6]The figure showing a sufferer's from rheumatism IgG neutral sugar chain analysis result.

[Drawing 7]The distribution map of sugar chain peak ratio P/H to a sufferer from rheumatism and a healthy person.

[Drawing 8]The lineblock diagram of the neural network for rheumatoid arthritis patient discernment.

[Drawing 9]The figure showing the contribution of each sugar chain in a rheumatoid arthritis patient's discernment.

[Drawing 10]The distribution map of sugar chain peak ratio G/F to an atopic dermatitis patient and a healthy person.

[Drawing 11]The distribution map of a rate occupied at the whole sugar chain peak F to a diabetic and a healthy person.

[Drawing 12]The distribution map of sugar chain peak ratio $(M+N+O+P)/(E+F+G+H)$ to a patient with liver disease and a healthy person.

[Translation done.]

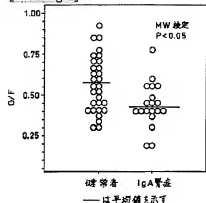
* NOTICES *

JPO and INPIT are not responsible for any damages caused by the use of this translation.

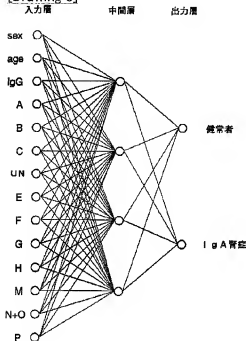
- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DRAWINGS

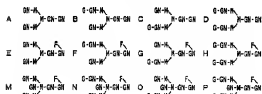
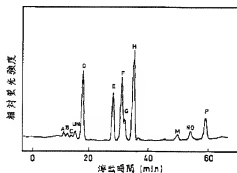
[Drawing 2]



[Drawing 3]

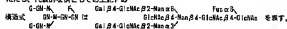


[Drawing 1]



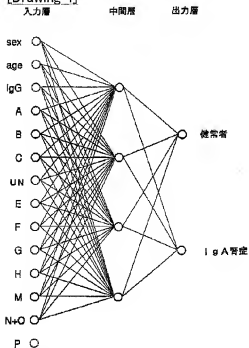
URL: ukopen.com 標酒表酒家

ただし、代表的な例としての上記Pの

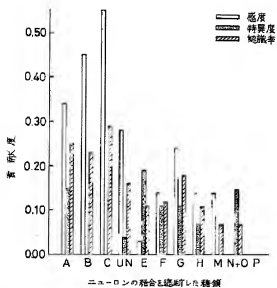


G, Gal:ガラクトース、M, Man:マンノース、F, Fuc:フコース、OH, Glucitol:D-アセチルグルコサミンを誘導。

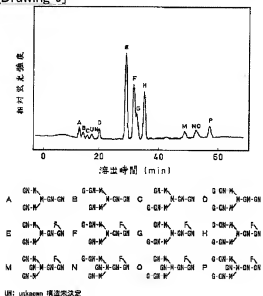
[Drawing 4]



[Drawing 5]



[Drawing 6]

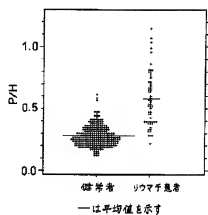


ただし、代表例としての上記Pの

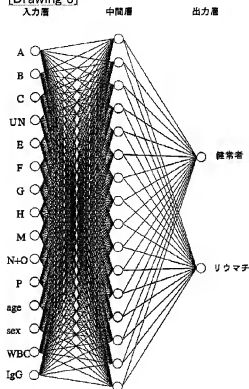
構造式 $\text{Glc}(\alpha\text{-Glc})\text{Glc}$ は $\text{Glc}(\beta\text{-Glc})\text{Glc}$ を表す。

5.Gal:ガラクトース、N.Nam:マンノース、F.Fuc:フコース、Glc,SiMe:シアセラルグルコサミンを表す。

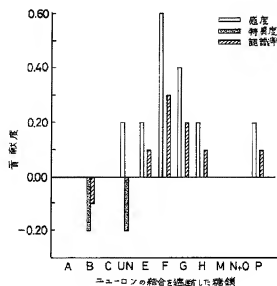
[Drawing 7]



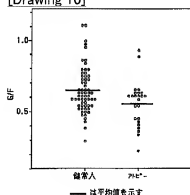
[Drawing 8]



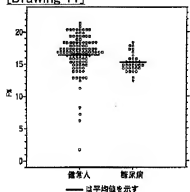
[Drawing 9]



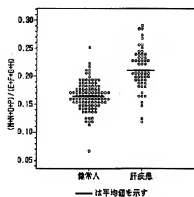
[Drawing 10]



[Drawing 11]



[Drawing 12]



[Translation done.]

特開平9-257790

(43) 公開日 平成9年(1997)10月3日

(51) Int.Cl. ⁶	識別記号	序内整理番号	F I	技術表示箇所
G 0 1 N	33/48		G 0 1 N	33/48 Z
	33/50			T
	33/66			Z

審査請求 未請求 請求項の数 8 O L (全 12 頁)

(21) 出願番号	特願平9-1921	(71) 出願人	390022644 株式会社中野酢店 愛知県半田市市中村町2丁目6番地
(22) 出願日	平成9年(1997)1月9日	(71) 出願人	592069403 岩田 彰 愛知県名古屋市中区松原3丁目3-18
(31) 優先権主張番号	特願平8-7426	(72) 発明者	山田 恵美子 愛知県高浜市呉竹町2-7-1 ニュート ピア21 B-101
(32) 優先日	平8(1996)1月19日	(72) 発明者	中川 裕章 愛知県名古屋市中区瑞穂区明前町10-28
(33) 優先権主張国	日本 (J P)	(74) 代理人	弁理士 平木 祐輔 (外1名)

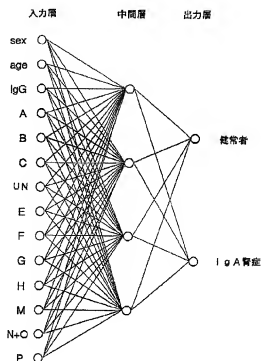
最終頁に続く

(54) 【発明の名称】 診断支援方法

(57) 【要約】

【課題】 組織及び組織液中の糖蛋白質や糖脂質に含まれる複数の糖鎖含有量を用いて精度の高い疾患その他の診断支援を行うことのできる方法を提供する。

【解決手段】 生体より採取した複数の糖鎖の濃度に相関するデータ A, B, ..., P をニューラルネットワークで処理して診断に有用な情報を得る。糖鎖の濃度に相関するデータとは、糖鎖の濃度あるいは濃度比、HPLC によって検出されたピークの面積あるいは面積比等糖鎖の濃度に直接的あるいは間接的に関連するデータである。ニューラルネットワークには、年齢、性別等診断に必要なそれ以外の臨床データを同時に入力するようにしてもよい。糖鎖を Ig G から得た糖鎖とすると、Ig A 腎症、リウマチ、アトピー性皮膚炎、糖尿病、肝疾患の診断支援を行うことができる。



【特許請求の範囲】

【請求項1】 生体より採取した複数の糖鎖の濃度に相関するデータをニューラルネットワークで処理して診断支援結果を得ることを特徴とする診断支援方法。

【請求項2】 前記糖鎖の濃度に相関するデータ以外の臨床データを前記ニューラルネットワークで同時に処理することを特徴とする請求項1記載の診断支援方法。

【請求項3】 前記糖鎖の濃度に相関するデータは、欠落データ及び飛び離れた値に各々専用の値を割り当てて前記ニューラルネットワークに入力することを特徴とする請求項1又は2記載の診断支援方法。

【請求項4】 前記糖鎖の濃度に相関するデータは、a及びbを $0 < a < b < 1$ を満たす数とすると、欠落データa、飛び離れた値を1とし、その他のデータはaとbの間に分布するように正規化して前記ニューラルネットワークに入力することを特徴とする請求項1又は2記載の診断支援方法。

【請求項5】 前記糖鎖はI g Gから得た糖鎖であり、I g A腎症又はリウマチの診断を行うことを特徴とする請求項1〜4のいずれか1項記載の診断支援方法。

【請求項6】 前記糖鎖はI g Gから得た糖鎖であり、アトピー性皮膚炎、糖尿病、又は肝炎、肝硬変もしくは肝細胞癌となる肝疾患の診断を行うことを特徴とする請求項1〜4のいずれか1項記載の診断支援方法。

【請求項7】 前記糖鎖はI g Gから得た糖鎖であり、疾患の重症度又は病型を識別することを特徴とする請求項1〜4のいずれか1項記載の診断支援方法。

【請求項8】 前記糖鎖はI g Gから得た糖鎖であり、複数の疾患間より単一疾患を識別することを特徴とする請求項1〜4のいずれか1項記載の診断支援方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、臨床データに基づいて病態等の診断支援を行う診断支援方法に関する。

【0002】

【従来の技術】ヒトの組織及び組織液中には多数の糖蛋白質、糖脂質があって、それらに含まれる糖鎖が何十種類も存在する。例えば、血清中の糖蛋白質には、約40種類のN結合型糖鎖が含まれている(Nakagawa, H., et al., (1995) Anal. Biochem., 226, 130-138.)。また、血清の免疫グロブリンGの中性糖鎖は多量類あることが知られている(Mizuchi, T., et al., (1982) J. Immunol., 129, 2016.)。これらの糖鎖の含有量を測定する方法としては、高速液体クロマトグラフィー(HPLC)を用いる方法、レクチンや抗体などを用いる方法、免疫電気泳動を用いる方法等がある。

【0003】現在、様々な疾患で、組織及び組織液中の糖蛋白質、糖脂質に含まれる複数の糖鎖構造が変化していることが知られている。例えば、リウマチでは、血清中の免疫グロブリンG(I g G)の16種類の糖鎖構造

のうち数種類の糖鎖濃度が変化していることが知られ

(Thomas W. Rademacher, et al., (1988) Springer Sem in. Immunopathol., 10, 231-249.)、また、肝細胞癌では血清中のトランスフェリンに含まれる糖鎖構造が変化していることが分かっている(Yamashita, K., et al., (1989) J. Biol. Chem., 264, 2415-2423.)。実際に癌で、細胞表面の糖鎖構造が変化し、その糖鎖を抗原とするモノクローナル抗体が腫瘍マーカーとして既に利用されている。肺癌では、糖鎖抗原が80%という高い陽性率で腫瘍マーカーになっている(梅山ら：腫瘍, 3, 22-33, 1988)。

【0004】

【発明が解決しようとする課題】現在のところ、糖鎖を抗原とするモノクローナル抗体を用いた単一の糖鎖含有量の測定値が腫瘍マーカーとして実用化されているだけで、例えばHPLCによる検出法など他の測定法を用いた複数の糖鎖含有量の測定値による疾患診断支援は実用化されていない。

【0005】糖鎖含有量を疾患診断に利用するために、は、患者と健康者の組織及び組織液中の糖蛋白質、糖脂質の糖鎖含有量を測定し、健康者と患者の糖鎖含有量の比較を行う必要がある。通常、この比較は、カットオフ値の設定をして患者の陽性率を計算する統計学的手法によって行われている。

【0006】この統計学的手法では、糖鎖含有量を糖鎖の種類毎に1種類ずつ全て解析し、それぞれを健康者と患者と比較するか、バイセクトード-アセチルグルコサミンを持つ糖鎖と持たない糖鎖等の何種類かの糖鎖の含有量を組み合わせて健康者と患者と比較する。また、生体内の糖鎖には、その含有量に年齢による差異や性差があるものも存在するため、それらも考慮に入れて総合的に患者と健康者との比較を行い、疾患診断に利用しなければならない。しかし、現在の統計学的手法では、糖鎖の種類が多いため糖鎖の種類を組み合わせてカットオフ値の設定などの操作が煩雑である上に、年齢や性別等も考慮して充分な解析を行うことは困難である。

【0007】このように、糖鎖が疾患等と相関を有することは明らかであるが、現在用いられている統計学的手法では、組織及び組織液中の糖蛋白質、糖脂質に含まれる複数の糖鎖含有量の総合的な解析が困難であるという問題がある。

【0008】本発明は、このような現状に鑑みてなされたもので、組織及び組織液中の糖蛋白質や糖脂質に含まれる複数の糖鎖含有量を用いて精度の高い疾患その他の診断を行うことのできる方法を提供することを目的とする。

【0009】

【課題を解決するための手段】本発明においては、生体より採取した複数の糖鎖の濃度に相関するデータをニューラルネットワークで処理して診断に有用な情報を得る

ことで前記目的を達成する。

【0010】糖鎖の濃度に相関するデータとは、例えば糖鎖の濃度あるいは濃度比、HPLCによって検出されたピークの面積あるいは面積比、レクチンや抗体により検出した凝集、発色や発光等の強度、免疫電気泳動で検出した染色濃度等、糖鎖の濃度に直接的あるいは間接的に関連するデータのことを意味する。

【0011】ニューラルネットワークには、糖鎖の濃度に相関するデータとともに診断に必要なそれ以外の年齢、性別、白血球数(WBC)、I g G濃度等の臨床データを適宜選択して同時に入力するようにしてもよい。

【0012】糖鎖の濃度に相関するデータに欠落データあるいは飛び離れ値(他の値に比べて極端に大きな値や小さな値のデータ)がある場合、それらのデータには各々専用の値を割り当ててニューラルネットワークに入力するのが好ましい。例えば、糖鎖の濃度に相関するデータは、 a 及び b を $0 < a < b < 1$ を満たす数とすると、 a 、欠落データを 0 、飛び離れ値を 1 とし、その他のデータは a と b の間に分布するように正規化して前記ニューラルネットワークに入力する。 a 、 b の値の一例を示すと、 $a = 0.1$ 、 $b = 0.9$ である。

【0013】糖鎖の濃度に相関するデータをI g Gから得た糖鎖データとすると、I g A腎症、慢性関節リウマチ、結核病、肝炎患、またはアトピー性皮膚炎の診断に有用な情報を得ることができる。

【0014】ニューラルネットワークは、高速な並列演算能力と学習機能を兼ね備え、多数の情報の解析に威力を発揮する。ニューラルネットワークとしては、例えば階層型ネットワークや相互結合型ネットワーク等を適宜採用することができ、その学習法としては、教師パターンとして与えられた出力パターンと実際の出力層のニューロンの出力値との誤差を逐次減少させるように結合荷重の変更を行い最適な結合荷重を決定する周知の誤差逆伝搬学習法等を採用することができる。

【0015】

【発明の実施の形態】以下、本発明の実施の形態について説明する。

【実施の形態1】I g A腎症の診断支援

複数のI g G中性糖鎖のデータを用いてI g A腎症の識別試験を行った。I g G中性糖鎖構造は、年齢(Raj, P 40 arekh, et al., (1988) Brief Definitive Report, 167, 1731-1736.)、性別(N. Tsuchiya, et al., (1993) J. Immun., 151, 1137-1146.)によって変化する事が知られているので、入力パターンとして年齢、性別、I g G濃度及びI g G中性糖鎖のピーク面積比(百分率)のデータを用いた。識別試験用のデータには、確定診断のついているI g A腎症患者19例と健康者30例のものを用いた。

【0016】I g G中性糖鎖は、次のようにして分析した。まず、血清からプロテインGを用いてI g Gを調製

した。調製したI g Gから酵素消化により糖鎖を切り出すと同時に酵素消化により全糖鎖を中性糖鎖とした。切り出した中性糖鎖を蛍光標識化し、オクタデシルシリル(octadecylsilyl、以下ODSと記す)カラム(Nakano pak ODS-A: (株) アジノキ製)を用いたHPLCにより12種類の中性糖鎖を分析した。ODSカラムによる分析結果と各ピークA~Pの糖鎖構造を図1に示す。

【0017】ピーク面積比(百分率)を求める際、ピークDは検体によってI g A腎症患者、健康者にかかわらずI g G以外の血清糖タンパク質由来の糖鎖が流入することがあるため、入力データからは除外した。また、ピークNとピークOは本来2つのピークに分離するが、カラムにより1つのピークとして検出されることがあるため、あらかじめN+Oの1つのピークとして計算した。ピークD以外の検出できた糖鎖ピーク11種類全ての面積に対して各ピーク面積の百分率を求めた。

【0018】最初に、比較のために、前記のようにして求めたI g Gの中性糖鎖ピーク面積比を用いて、従来の統計学的Mann-Whitneyの有意差検定法によりI g A腎症の識別を行うことを検討した。その結果、図2に示す糖鎖ピークFとGの比率にのみ危険率5%以下で有意差が認められた。このとき、カットオフ値を平均値±(2×標準偏差)に設定した場合のI g A腎症患者と健康者との認識率は65.3%、特異度は100%、感度は10.5%であった。ただし、認識率、特異度及び感度は、患者検体数をNP、健康者検体数をNH、患者を患者として識別した検体数をnp、健康者を健康者として識別した検体数をnhとすると、それぞれ次式で定義される。

【0019】認識率 $= (np + nh) / (NP + NH)$

感度 $= np / NP$

特異度 $= nh / NH$

【0020】次に、ニューラルネットワークを用いる本発明の方法について説明する。ここでは、図1に示すピークD以外の検出できた糖鎖ピーク11種類全ての面積に対して各ピーク面積の百分率を求め、入力データとした。検出したピークを全て入力パターンとして用いることにしたのは、以下の理由による。今まで、疾患によって変化すると言われている糖鎖構造は、顕著な濃度の変化や、統計上の単純な糖鎖の組み合わせによる統計値の変化によってしかとらえられていない。しかし、体内に多くの種類の糖鎖が存在するということを考えると、今までに行われた解析では糖鎖が多少情報をも十分に引き出していない可能性が残るため、ニューラルネットワークによる複雑なパターン解析で全ての糖鎖についての解析を行うことにしたものである。

【0021】また、前記I g G11種類の糖鎖ピークからさらに図1に示す糖鎖ピークのうちA、B、C、UNの4種類のデータを除いて、残り7種類に対して各糖鎖ピーク面積の百分率を求め、入力データとして用いることがで

きる。ODSカラムを用いたHPLCにより糖鎖ピークA、B、C、UNが検出される溶出位置には、糖鎖の蛍光標識の際に生じる反応副成分が重なって検出されるため、糖鎖ピークA、B、C、UNを除外するとデータの反応条件によるばらつきを抑えることができる。

【0022】このような7種類のI g G中性糖鎖のピーク面積比を入力データとして用い、複数疾患から単一疾患の識別実験を行ったとき、感度、認識率がわずかに向上した。ニューラルネットワークとしては、図3に示すような3層の階層型ネットワークを採用した。

【0023】ここでは、入力パターンとして、年齢、性別、I g G濃度及び前記11種類のI g G中性糖鎖のピーク面積比(百分率)のデータ、すなわち全部で14項目のデータを用いて2種類の出力、すなわちI g A腎症であるか否かを識別するので、入力層のニューロン数は14個、出力層のニューロン数は2個とした。出力層の2個のニューロンは出力値0(健康者である)及び1(I g A腎症である)に対応している。この例で出力層のニューロン数を2個としたのは単に識別の有意性を明らかにするためであり、出力層のニューロン数は1個であってもよいし、3個以上としてI g A腎症である可能性の確率あるいは危険率を表示するようにしてもよい。

【0024】中間層のニューロン数は、少なすぎると入力パターンの複雑な組み合わせが出力層に伝わらず学習が収束しないし、また逆に多すぎると学習回数が多くなり時間がかかって煩雑になる。そのため、中間層のニューロン数は、実験的に最適数を決めた。ここでは、中間層のニューロン数を1から30まで変化させて誤差逆伝搬法により学習させ、識別実験を行った中で最も成績が良かった場合の中間層のニューロン数、4個を採用した。

【0025】ニューラルネットワークの学習は、前述の確定診断のついてる健康者30例のデータとI g A腎症患者19例のデータから11例をランダムに取り出して30例としたI g A腎症患者30例、合計60例のデータを用い、希望する出力パターンを教師パターンとして与える誤差逆伝搬学習法によった。まず、年齢、I g G濃度、上記分析法によるI g G中性糖鎖11種類の各ピーク面積比の各々を次の(数1)により正規化した。ただし、欠落データに対しては $a=0$ として入力すること、欠落情報に対処可能なニューラルネットワークとし、また他の値に比べて極端に大きなデータや小さなデータ(飛び離れた値)に対しては $a=1$ とした。性別に関しては男性を0.1、女性を0.9として数値化した。飛び離れた値は、例えばスミルノフの棄却検定により棄却できる値とする(市原清志「バイオサイエンスの統計学」南江堂出版(1990年)第284～285頁)、あるいは経験的に任意な値に決定してもよい。ここでは、スミルノフの棄却検定により棄却できる値は全て $a=1$ とし、棄却できない値は(数1)により正規化し、0、

1～0.9の間に数値化して入力した。

【0026】

【数1】 $a = \{ (a_0 - a_{min}) / (a_{max} - a_{min}) \} \times 0.8 + 0.1$

a : 正規化後のデータ

a_0 : 正規化前のデータ

a_{max} : 最大値

a_{min} : 最小値

【0027】入力値は、それぞれ結合荷重を乗ぜられて中間層に伝わり、結合荷重が乗ぜられた各入力値は加算され、応答関数による変形を受けて出力される。中間層からの出力に、また結合荷重が乗ぜられて出力層において加算され、応答関数による変形を受けて出力される。応答関数としてはシグモイド関数を用いた。誤差逆伝搬学習法により、与えられた教師パターンと実際の出力層の値との誤差を逐次減少させるように結合荷重の変更を行うことを全ての教師パターンに対して繰り返すことにより、結合荷重を最適な値に変化させる。

【0028】前記したI g A腎症患者と健康者各30例(合計60例)のデータを用い、リーブ・ワン・アウト(leave one out)法、すなわち59例のデータで学習終了したニューラルネットワークを用いて残りの1例について識別を行う事を全60例のデータについて各々繰り返して行う方法でI g A腎症患者の認識率及び感度を調べたところ、認識率が93.3%、感度が96.7%と比較例で説明した従来の統計学的手法に比べていづれも飛躍的に向上した。

【0029】次に、各糖鎖データがどのようにI g A腎症の識別に貢献しているかを次のようにして見出した。

図4に示すように、11種類のI g G中性糖鎖ピーク面積比のうち任意の1つ(図4の例ではピークP)の入力層と中間層との結合を遮断したときの認識率の変化を11種類の全ての糖鎖について調べた。糖鎖ピーク面積比のデータの出力層と中間層とを全て結合した場合の認識率を x_0 、ある1種類の糖鎖ピーク面積比のデータの出力層と中間層との結合を遮断したときの認識率を x とするとき、その入力層と中間層との結合を遮断した糖鎖の認識率に対する貢献度 y を次の(数2)で定義する。

【0030】

【数2】 $y = (x_0 - x) / x_0$

【0031】特異度及び感度に関しても同様にして貢献度を定義する。図5は、I g A腎症患者の識別における各糖鎖データの貢献度を認識率、特異度、感度の各々に対してグラフ化して示したものである。大きな貢献度を示している糖鎖ほどニューラルネットワークによるI g A腎症の識別に対して有用であることを意味している。この場合、図5から糖鎖ピークP以外の全ての糖鎖データは、I g A腎症の識別に関して効果的に作用しており、これらの糖鎖データの複合的な組み合わせによりニューラルネットワークが識別を行い、認識率を向上させ

ていることが分る。

【0032】従来の統計的手法による識別結果、全てのニューロンを結合させたニューラルネットワークによる識別結果を表1にまとめて示す。

【0033】

【表1】

I g G抗体の識別結果 (感度%)

	感 度	特異度	認識率
統計的手法	10.5	100	65.3
ニューラルネットワーク (本発明)	85.7	90.0	88.3

【0034】このように、I g G中性糖鎖ピーク面積比データを解析する手段としてニューラルネットワークを用いることにより、従来の統計的手法では充分に行うことが出来なかった多種類の糖鎖情報を組み合わせた解析を自動的に行うことができ、しかもその解析による疾患と非疾患との認識率の向上を図ることが出来る。

【0035】ここではODSカラムを用いたHPLCにより糖鎖を分析し、各ピーク面積比のデータを入力パターンとして採用する例を説明したが、糖鎖の検出は他のカラム (ゲル濾過、アフィニティークラムなど) を用いたり、レクチンや抗体を用いる方法、免疫電気泳動法などによってもよいし、これらの方法で測定した糖鎖の濃度などのデータを入力パターンとして用いてもよい。

【0036】〔実施の形態2〕慢性関節リウマチの診断支援

年齢、性別、I g G濃度、WBC (白血球) 数、ODSカラムを用いたHPLCで分析したI g G中性糖鎖のピーク面積比 (百分率) のデータを入力パターンとして用いて、慢性関節リウマチの識別試験を行った。識別試験用のデータとしては、確定診断のついているリウマチ患者59例と健康者278例のものを用いた。

【0037】I g G中性糖鎖の分析は実施の形態1と同様に行った。ODSカラムによる分析結果と各ピークA～Pの糖鎖構造を図6に示す。最初に、比較のために、I g Gの中性糖鎖ピーク面積比を用いて、従来の統計学的 Mann-Whitney の有意差検定法により慢性関節リウマチの識別を行うことを検討した。その結果、図7に示すように、危険率0.1%以下で有意差が認められた糖鎖ピークHとPの比率で、カットオフ値を平均値±(2×標準偏差)に設定した場合のリウマチ患者と健康者の認識率は91.4%、特異度は96.8%、感度は66.1%であった。

【0038】次に、ニューラルネットワークを用いる識別方法について説明する。ニューラルネットワークは、図8に示すように、実施の形態1と同様なシグモイド関数を用いる3層の階層型ネットワークとした。15項目の糖鎖及び臨床上の情報に基づいてリウマチであるか否かを識別するので、入力層のニューロン数は15個、出

力層は2個とした。出力層の2個のニューロンは出力値0 (健康である)、1 (リウマチである) に対応している。中間層のニューロン数は実験により最適数を決めた。ここでは、中間層のニューロン数を1～30まで変化させて誤差逆伝搬法により学習させ、最も効率よく学習を行えた場合の中間層のニューロン数を採用した。この場合は16個とした。

【0039】ニューラルネットワークの学習は、前述の確定診断のついている慢性関節リウマチ患者と健康者各50例のデータを用い、希望する出力パターンを教師パターンとして与える誤差逆伝搬学習法によった。今回は、実用上の制限がある場合を想定し、設定した誤差率に最短時間に到達したネットワークを用いた。まず、年齢、I g G濃度、WBC数、上記分析法によるI g G中性糖鎖11種類の糖鎖ピーク面積比の各々に実施の形態1と同様に前記(数1)による正規化処理を施し、入力パターンとした。ただし、欠落データに対してはa=0として欠落情報に対処可能なネットワークとし、(また他の値に比べて極端に大きなデータや小さなデータ (飛び離れた値) に対してはa=1とした。性別に関しては男性を0.1、女性を0.9として数値化した。

【0040】前記した慢性関節リウマチ患者と健康者各50例 (合計100例) のデータを用い、リーブ・ワン・アウト (leave one out) 法、すなわち99例のデータで学習終了したニューラルネットワークを用いて残りの1例について識別を行う事を全100例のデータについて各々繰り返す方法で慢性関節リウマチ患者の認識率を調べた。その結果、認識率96.0%、特異度96.0%、感度96.0%となり感度が向上した。

【0041】次に、実施の形態1と同様に慢性関節リウマチの識別に各糖鎖データがどのように貢献しているかを見出した。図9は、図5と同様にして慢性関節リウマチ患者の識別における各糖鎖の貢献度をグラフ化して示したものである。貢献度の大きい糖鎖ほどニューラルネットワークによる慢性関節リウマチ患者の識別に対して有用である。

【0042】図9の結果を参照して、ニューラルネットワークによる慢性関節リウマチ患者の識別に貢献度の低い糖鎖のデータとして糖鎖ピークBのデータの入力層と中間層との結合を遮断したところ、特異度が98.0%に向上し、認識率が97.0%に向上した。

【0043】従来の統計的手法による識別結果、糖鎖のデータの入力層と中間層とを全て結合した場合のニューラルネットワークによる識別結果、及び糖鎖ピークBのデータの入力層と中間層との結合を遮断したニューラルネットワークによる識別結果を表2にまとめて示す。

【0044】

【表2】

慢性関節リウマチの識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	66.1	96.8	91.4
全てのニューロンを結合したニューラルネットワーク (本発明)	96.0	96.0	96.0
貢獻度の低いニューロンとの結合を遮断したニューラルネットワーク (本発明)	96.0	98.0	97.0

【0045】〔実施の形態3〕アトピー性皮膚炎の診断支援

年齢、性別、ODSカラムを用いたHPLCで分析したI g G中性糖鎖のピーク面積比(百分率)のデータを入力パターンとして用いて、アトピー性皮膚炎の識別試験を行った。識別試験用のデータとしては、確定診断のついているアトピー性皮膚炎患者24例と健常者63例のものをを用いた。

【0046】まず、比較のために、従来の統計学的 Mann-Whitney の有意差検定法によりアトピー性皮膚炎の識別を行うことを検討したところ、図10に示すように、糖鎖ピークDとGの比率に危険率5%以下で有意差が認められた。このとき、カットオフ値を平均値土(2×標準偏差)に設定した場合のアトピー性皮膚炎患者と健常者の認識率は69.0%、特異度は92.1%、感度は8.3%であった。

【0047】次に、ニューラルネットワークを用いる識別方法について説明する。ニューラルネットワークの入力層のニューロン数は、年齢、性別、及びI g G糖鎖11種類の13個とし、出力層のニューロン数はアトピー性皮膚炎と健常の2個とした。中間層のニューロン数

*30
アトピー性皮膚炎の識別結果 (数値は%)

	感 度	特異度	認識率
統計学的手法	8.3	92.1	69.0
ニューラルネットワーク (本発明)	89.4	83.3	86.3

【0051】〔実施の形態4〕糖尿病の診断支援

年齢、性別、ODSカラムを用いたHPLCで分析したI g G中性糖鎖のピーク面積比(百分率)のデータを入力パターンとして用いて、糖尿病の識別試験を行った。識別試験用のデータとしては、確定診断のついている糖尿病患者34例と健常者110例のものをを用いた。

【0052】まず、比較のために、従来の統計学的 Mann-Whitney の有意差検定法により糖尿病の識別を行うことを検討したところ、図11に示すように、糖鎖ピークFの割合に危険率0.1%以下で有意差が認められた。このとき、カットオフ値を平均値土(2×標準偏差)に設定した場合の糖尿病患者と健常者の認識率は73.6%、特異度は95.5%、感度は2.9%であった。

【0053】次に、ニューラルネットワークを用いる識

10*は、経験則に基づいて決定し4個とした。年齢、性別、I g G糖鎖の各データは実施の形態1と同様にして正規化処理を施して入力パターンとした。ニューラルネットワークの学習は、前述の確定診断のついているアトピー性皮膚炎患者24例と、健常者63例の中から無作為に選んだ24例のデータとを用い、希望する出力パターンを教師パターンとして与える誤差逆伝播学習法によった。

【0048】このアトピー性皮膚炎患者と健常者各24例(合計48例)のデータを用い、リーブ・ワン・アウト法によりアトピー性皮膚炎患者の認識率を調べたところ、感度89.4%、特異度83.3%、認識率86.3%が得られた。このようにニューラルネットワークを用いることで、従来の統計学的手法に比べて感度、認識率が向上した。

【0049】従来の統計学的手法による識別結果と、ニューラルネットワークによる識別結果を表3にまとめて示す。

【0050】

【表3】

別方法について説明する。ニューラルネットワークの入力層のニューロン数は、年齢、性別、及びI g G糖鎖11種類の13個とし、出力層のニューロン数は糖尿病患者34例と、健常者110例の中から無作為に選んだ34例のデータとを用い、希望する出力パターンを教師パターンとして与える誤差逆伝播学習法によった。

【0054】この糖尿病患者と健常者各34例(合計68例)のデータを用い、リーブ・ワン・アウト法により糖尿病患者の認識率を調べたところ、感度100%、特

異度100%、認識率100%が得られた。このようにニューラルネットワークを用いることで、従来の統計学的手法に比べて感度、特異度、認識率が向上した。

【0055】従来の統計学的手法による識別結果と、ニセ増原病の識別結果（数値は%）

	感 度	特異度	認識率
統計学的手法	2.9	95.5	73.6
ニューラルネットワーク（本発明）	100	100	100

【0057】【実施の形態5】肝疾患の診断支援
年齢、性別、ODSカラムを用いたHPLCで分析したI g G中性糖鎖のピーク面積比（百分率）のデータを入力パターンとして用いて、肝疾患の識別試験を行った。識別試験用のデータとしては、確定診断のついている肝疾患患者60例（肝炎20例、肝硬変20例、肝細胞癌20例）と健常者110例のものを用いた。

【0058】まず、比較のために、従来の統計学的 Mann-Whitney の有意差検定法により肝疾患の識別を行うことを検討したところ、図12に示すように、バイセクトン-Nアセチルグルコサミンを持つ糖鎖と持たない糖鎖の比 $(M+N+O+P) / (E+F+G+H)$ に危険率0.1%以下で有意差が認められた。このとき、カットオフ値を平均値 $\pm (2 \times \text{標準偏差})$ に設定した場合の肝疾患患者と健常者との認識率は75.9%、特異度は93.6%、感度は43.3%であった。

【0059】次に、ニューラルネットワークを用いる識別方法について説明する。ニューラルネットワークの入力層のニューロン数は、年齢、性別、及びI g G糖鎖11種類の13個とし、出力層のニューロン数は肝疾患とニセ増原病の識別結果（数値は%）

	感 度	特異度	認識率
統計学的手法	43.3	93.6	75.9
ニューラルネットワーク（本発明）	87.8	83.9	85.9

【0063】【実施の形態6】慢性関節リウマチの診断支援（重症度識別）

慢性関節リウマチ患者と健常者のデータを用い、ニューラルネットワークにより重症度の識別を行った。入力パターンとして用いたデータは実施の形態2と同じである。

【0064】識別試験用のデータとして、確定診断の付いている慢性関節リウマチ患者48例と健常者282例を用いた。慢性関節リウマチ患者は、関節の器質的変化に基づく重症度としてステージ1〜4に分けられており、今の場合、ステージ1の患者が15例、ステージ2の患者が11例、ステージ3の患者が6例、ステージ4の患者が16例であることを、無作為に選んだデータを加えて各ステージ20例とした。重症度の分類は Steinbr

＊ニューラルネットワークによる識別結果を表4にまとめて示す。

【0056】

【表4】

※健常の2個とした。中間層のニューロン数は、経験則に基づいて決定し4個とした。年齢、性別、I g G糖鎖の各データは実施の形態1と同様にして正規化処理を施して入力パターンとした。ニューラルネットワークの学習は、前述の確定診断のついている肝疾患患者60例と、健常者110例の中から無作為に選んだ60例のデータとを用い、希望する出力パターンを教師パターンとして与える誤差逆伝搬学習法によった。

【0060】この群尿病患者と健常者各60例（合計120例）のデータを用い、リーブ・ワン・アウト法により肝疾患患者の認識率を調べたところ、感度87.8%、特異度83.9%、認識率85.9%が得られた。このようにニューラルネットワークを用いることで、従来の統計学的手法に比べて感度、認識率が向上した。

【0061】従来の統計学的手法による識別結果と、ニューラルネットワークによる識別結果を表5にまとめて示す。

【0062】

【表5】

ocker のステージ分類による（本邦臨床統計集、512-521(1993)日本臨牀社）。健常者は、282例から20例を無作為に選んだ。

【0065】ニューラルネットワークの入力層のニューロン数は、実施の形態2と同様に15個とし、出力層のニューロン数は各ステージと健常の5個とした。中間層のニューロン数は、経験則に基づいて決定し6個とした。年齢、性別、I g G濃度、WBC数、I g G糖鎖の各データは実施の形態1と同様にして正規化処理を施して入力パターンとした。ニューラルネットワークの学習は、前述の確定診断のついているステージ1〜ステージ4の慢性関節リウマチ患者各20例と、健常者282例の中から無作為に選んだ20例のデータとを用い、希望する出力パターンを教師パターンとして与える誤差逆伝

機学習法によった。

【0066】この慢性関節リウマチ患者と健康者のデータを用い、リーブ・ワン・アウト法により慢性関節リウマチの重症度を識別したところ、表6に示すように、感度はステージ1が84.5%、ステージ2が97.5%、ステージ3が100.0%、ステージ4が87.5%であ

慢性関節リウマチの重症度識別結果

平 均	感 度 (%)				特異度 (%)	認識率 (%)
	ステージ1	ステージ2	ステージ3	ステージ4		
82.4	84.5	97.5	100.0	87.5	96.0	93.1

【0068】〔実施の形態7〕糖尿病の診断支援（病型識別）

糖尿病患者と健康者のデータを用い、ニューラルネットワークにより糖尿病の病型識別を行った。入力パターンとして、年齢、性別、I g G糖鎖のデータを用いた。

【0069】識別試験用のデータとして、確定診断のついている糖尿病患者34例（うちI型5例、II型29例）と健康者110例を用いた。糖尿病の病型の分類は1980年のWHOの分類による（内科、75、1524-1527（1995）南江堂）。インスリンの不足が絶対的であるものはI型糖尿病、インスリンの不足が相対的であるものはII型糖尿病に分類される。I型糖尿病はインスリン依存型糖尿病、II型糖尿病はインスリン非依存型糖尿病ともいう。患者は、無作為に選んだデータを加えてI型、II型各30例とした。健康者は、110例から40例を無作為に選んだ。

【0070】ニューラルネットワークの入力層のニューロン数の病型識別結果

平 均	感 度 (%)		特異度 (%)	認識率 (%)
	I 型	II 型		
96.7	100.0	93.2	100.0	98.0

【0073】〔実施の形態8〕複数疾患から単一疾患の識別

各疾患毎の患者のデータと健康者のデータを用い、ニューラルネットワークにより単一疾患の識別を行った。入力パターンとして、年齢、性別、I g G糖鎖のデータを用いた。

【0074】識別試験用のデータとして、それぞれ確定診断のついているI g A腎症患者19例と健康者30例、慢性関節リウマチ患者48例と健康者282例、アトピー性皮膚炎患者24例と健康者63例、糖尿病患者34例と健康者110例、肝疾患患者60例と健康者110例、以上のデータから無作為にそれぞれ患者を30例、健康者を6例ずつ取り出し、全体で、I g A腎症患者30例、慢性関節リウマチ患者30例、アトピー性皮膚炎患者30例、糖尿病患者30例、肝疾患患者30

り、平均92.4%であった。また、特異度96.0%、認識率93.1%が得られた。このようにニューラルネットワークを用いることで、慢性関節リウマチの重症度を識別することができた。

【0071】

【表6】

※ロン数は、年齢、性別、I g G糖鎖11種類の13個とし、出力層のニューロン数は糖尿病I型、糖尿病II型、健康の3個とした。中間層のニューロン数は、経験則に基づいて決定し4個とした。各データは実施の形態1と同様にして正規化処理を施して入力パターンとした。ニューラルネットワークの学習は、希望する出力パターンを教師パターンとして与える誤差逆伝搬学習法によった。

【0071】この糖尿病患者と健康者のデータを用い、リーブ・ワン・アウト法により糖尿病の病型を識別したところ、表7に示すように、感度は糖尿病I型が100.0%、糖尿病II型が93.2%で、平均96.7%が得られた。また、特異度100.0%、認識率98.0%が得られた。このようにニューラルネットワークを用いることで、糖尿病の病型を識別することができた。

【0072】

【表7】

例、健康者30例の、計180例として識別試験を行った。ただし、アトピー性皮膚炎患者とI g A腎症患者については、各々の患者の中から無作為に選出されたデータを加えて30例とした。

【0075】ニューラルネットワークの入力層のニューロン数は、年齢、性別、I g G糖鎖11種類の13個とし、出力層のニューロン数はI g A腎症患者、慢性関節リウマチ患者、アトピー性皮膚炎患者、糖尿病患者、肝疾患患者、健康者の6個とした。中間層のニューロン数は、経験則に基づいて決定し12個とした。各データは実施の形態1と同様にして正規化処理を施して入力パターンとした。ニューラルネットワークの学習は、希望する出力パターンを教師パターンとして与える誤差逆伝搬学習法によった。

【0076】前記複数疾患患者と健康者のデータを用

い、リーブ・ワン・アウト法により疾患を識別したところ、表8に示すように、感度はIgA腎症で90.7%、リウマチで81.7%、アトピー性皮膚炎で84.3%、糖尿病で87.0%、肝疾患で88.0%、平均で86.3%が得られた。また、特異度は61.3%、認識率※は82.2%であった。このようにニューラルネットワークを用いることで、複数疾患から単一疾患を識別することができた。

【0077】
【表8】

感 度 (%)					特異度	認識率
平均	IgA腎症	リウマチ	アトピー	糖尿病	(%)	(%)
86.3	90.7	81.7	84.3	87.0	61.3	82.2

【0078】

【発明の効果】本発明によると、糖鎖データを解析する手段としてニューラルネットワークを用いることにより、従来の統計的手法では充分に行うことが出来なかった多種類の糖鎖含有量を組み合わせた解析を自動的に行うことができる。

【図面の簡単な説明】

【図1】IgA腎症患者のIgG中性糖鎖分析結果を示す図。

【図2】IgA腎症と健常者に対する糖鎖ピーク比G/Fの分布図。

【図3】IgA腎症患者識別用ニューラルネットワークの構成図。

【図4】1つの糖鎖データについて、入力層と中間層との結合を遮断したニューラルネットワークの説明図。

【図5】IgA腎症患者の識別における各糖鎖データの※

※貢献度を示す図。

【図6】リウマチ患者のIgG中性糖鎖分析結果を示す図。

【図7】リウマチ患者と健常者に対する糖鎖ピーク比P/Hの分布図。

【図8】慢性関節リウマチ患者識別用ニューラルネットワークの構成図。

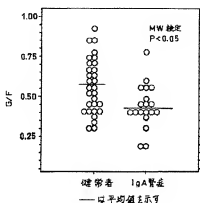
【図9】慢性関節リウマチ患者の識別における各糖鎖の貢献度を示す図。

【図10】アトピー性皮膚炎患者と健常者に対する糖鎖ピーク比G/Fの分布図。

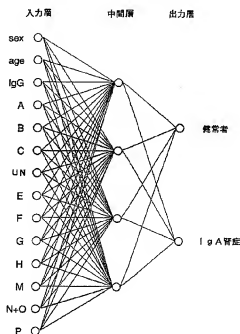
【図11】糖尿病患者と健常者に対する糖鎖ピークFの全体に占める割合の分布図。

【図12】肝疾患患者と健常者に対する糖鎖ピーク比(M+N+O+P)/(E+F+G+H)の分布図。

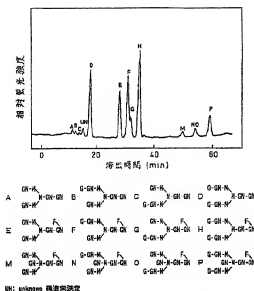
【図2】



【図3】



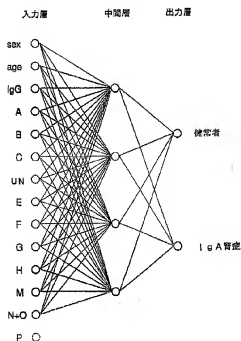
【図1】



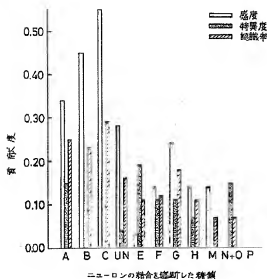
ただし、代表例としての上記の
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$ 、 $\text{Fuc}\alpha 6$
 $\text{GlcNAc}\beta 4\text{-Man}\beta 6\text{-GlcNAc}\beta 4\text{-GlcNAc}\beta 6\text{-GlcNAc}$ を表す。
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$

0-Gal:ガラクトース、M,Man:マンノース、F,Fuc:フコース、Glc,GlcNAc: N-アセチルグルコサミンを表す。

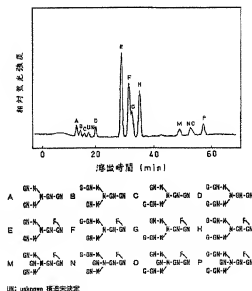
【図4】



【図5】



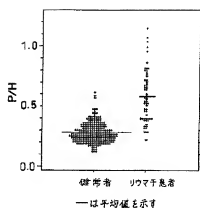
【図6】



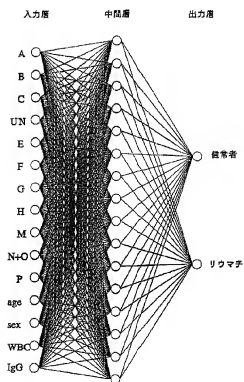
ただし、代表例としての上記の
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$ 、 $\text{Fuc}\alpha 6$
 $\text{GlcNAc}\beta 4\text{-Man}\beta 6\text{-GlcNAc}\beta 4\text{-GlcNAc}\beta 6\text{-GlcNAc}$ を表す。
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$
 O-Glc-Na は $\text{Gal}\beta 4\text{-GlcNAc}\beta 2\text{-Man}\alpha 6$

0-Gal:ガラクトース、M,Man:マンノース、F,Fuc:フコース、Glc,GlcNAc: N-アセチルグルコサミンを表す。

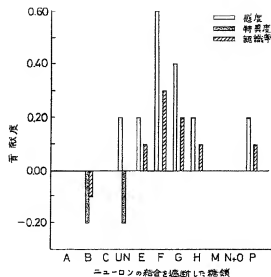
【図7】



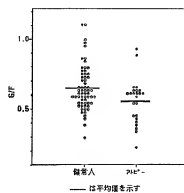
【図8】



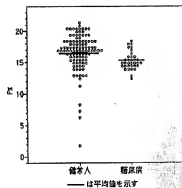
【図9】



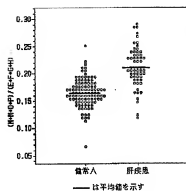
【図10】



【図 11】



【図 12】



フロントページの続き

(72)発明者 高橋 穂子
愛知県名古屋市長区松栄町 1-105
(72)発明者 塚本 義則
愛知県半田市清城町 3-3-23

(72)発明者 川村 吉也
愛知県江南市古知野町古渡132
(72)発明者 小栗 宏次
愛知県半田市中村町 1-18
(72)発明者 岩田 彰
愛知県名古屋市中区松原 3-3-18